

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

Claim 1. (Currently Amended) A computer system for generating data structures for information retrieval of documents stored in a database, said documents being stored as document-keyword vectors generated from a predetermined keyword list, and said document-keyword vectors forming nodes of a hierarchical structure imposed upon said documents, said computer system comprising:

a document-key word matrix generation subsystem;

a neighborhood patch generation subsystem for generating groups of nodes having similarities as determined using a search structure, said neighborhood patch generation subsystem including a subsystem for generating a spatial approximation sample hierarchy ~~hierarchical~~ structure upon said document-keyword vectors and a patch defining subsystem for creating patch relationships among said nodes with respect to a metric distance between nodes;

a query vector generation subsystem accepting search conditions and query keywords, generating a corresponding query vector, and storing the generated query vector;

[[a]] an intra-patch confidence and intrapath confidence determination subsystem for every element of the database, the spatial approximation sample hierarchy structure computing a neighborhood patch consisting of a list of those database elements most similar to it for computing intra-patch confidence values between patches and interpath confidence values; and

a self confidence determining subsystem for (a) computing a list of self confidence values, for every stored patch, (b) computing relative self confidence values, and (c)

thereafter using the relative self confidence values to determine a size of a best subset of each patch to serve as a cluster candidate;

a cluster estimation subsystem for generating cluster data of said document-keyword vectors using said similarities of patches wherein the cluster estimation subsystem selects said patches depending on ~~inner-patch~~ intra-patch confidence values to represent clusters of said document keyword vectors, estimate the sizes of said patches, and generate cluster data of document keyword vectors using similarities of the patches; and

a redundant cluster elimination subsystem for using the inner patch confidence values to eliminate redundant cluster candidates.

Claim 2. (Currently Amended) The computer system of claim 1, wherein said cluster estimation subsystem selects said patches depending on said ~~inter~~ inner-patch confidence values to represent clusters of said document-keyword vectors.

Claim 3. (Original) The computer system of claim 1, wherein said cluster estimation subsystem estimates sizes of said clusters depending on said intra-patch confidence values.

Claim 4. (Currently Amended) A method for generating data structures for information retrieval of documents stored in a database, said documents being stored as document-keyword vectors generated from a predetermined keyword list, and said document-keyword vectors forming nodes of a hierarchical structure imposed upon said documents, said method comprising the steps of:

generating a hierarchical structure upon said document-keyword vectors and storing hierarchy data in an adequate storage area;

accepting search conditions and query keywords, generating a corresponding query vector, and storing the generated query vector;

generating neighborhood patches of nodes having similarities as determined using levels of the hierarchical structure, and storing said patches in an adequate storage area;

generating groups of nodes having similarities as determined using a search structure, including generating a spatial approximation sample hierarchy structure upon said document-keyword vectors and creating patch relationships among said nodes with respect to a metric distance between nodes;

determining intra-patch confidence values between patches and interpatch confidence values

determining an intra-patch confidence and intrapath confidence for every element of the database, comprising utilizing the spatial approximation sample hierarchy structure to compute a neighborhood patch consisting of a list of those database elements most similar to it and computing intra-patch confidence values between patches and interpatch confidence values;

determining self confidence values to determine a size of a best subset of each patch to serve as a cluster candidate by the steps of (a) computing a list of self confidence values, for every stored patch, (b) computing relative self confidence values, and (c) thereafter using the relative self confidence values to determine the size of a best subset of each patch to serve as a cluster candidate;

invoking said hierarchy data and said patches to compute inter-patch confidence values between said patches and intra-patch confidence values, and storing said values as corresponding lists in an adequate storage area;~~and~~

estimating the sizes of said patches, and generating cluster data of document keyword vectors using similarities of the patches, selecting said patches depending on said inter-patch confidence values and said intra-patch confidence values to represent clusters of said document-keyword vectors: and

using the inner patch confidence values to eliminate redundant cluster candidates.

Claim 5. (Original) The method according to claim 4 further comprising the step of estimating sizes of said clusters depending on said intra-patch confidence values.

Claim 6 - 7. (Canceled)

Claim 8. (Currently Amended) A computer readable medium storing a program for making a computer system execute a method for generating data structures for information retrieval of documents stored in a database, said documents being stored as document-keyword vectors generated from a predetermined keyword list, and said document-keyword vectors forming nodes of a hierarchical structure imposed upon said documents, said program making said computer system execute the steps of:

accepting search conditions and query keywords, generating a corresponding query vector, and storing the generated query vector;

generating a hierarchical structure upon said document-keyword vectors and storing hierarchy data in an adequate storage area;

generating neighborhood patches consisting of nodes having similarities as determined using levels of the hierarchical structure, and storing said patch list in an adequate storage area;

generating groups of nodes having similarities as determined using a search structure, including generating a spatial approximation sample hierarchy structure upon said document-keyword vectors and creating patch relationships among said nodes with respect to a metric distance between nodes;

determining an intra-patch confidence and intrapath confidence for every element of the database, comprising utilizing the spatial approximation sample hierarchy structure to compute a neighborhood patch consisting of a list of those database elements most similar to it and computing intra-patch confidence values between patches and interpath confidence values;

determining self confidence values to determine a size of a best subset of each patch to serve as a cluster candidate by the steps of (a) computing a list of self confidence values, for every stored patch, (b) computing relative self confidence values, and (c) thereafter using the relative self confidence values to determine the size of a best subset of each patch to serve as a cluster candidate;

invoking said hierarchy data and said patches to compute inter-patch confidence values between said patches and intra-patch confidence values, and storing said values as corresponding lists in an adequate storage area;~~and~~

selecting said patches depending on said inter-patch confidence values and said intra-patch confidence values to represent clusters of said document-keyword vectors; and

using the inner patch confidence values to eliminate redundant cluster candidates.

Claim 9. (Original) The method according to claim 8, further comprising the step of estimating sizes of said clusters depending on said intra-patch confidence values.

Claim 10. (Withdrawn) An information retrieval system for of documents stored in a database, said documents being stored as document-keyword vectors generated from a predetermined keyword list, and said document-keyword vectors forming nodes of a hierarchical structure imposed upon said documents, said system comprising:

a neighborhood patch generation subsystem for generating groups of nodes having similarities as determined using a hierarchical structure, said patch generation subsystem including a subsystem for generating a hierarchical structure upon said document-keyword vectors and a patch defining subsystem for creating patch relationships among said nodes with respect to a metric distance between nodes; and

a cluster estimation subsystem for generating cluster data of said document-keyword vectors using said similarities of patches; and

a graphical user interface subsystem for presenting said estimated cluster data on a display means.

Claim 11. (Withdrawn) The computer system of claim 10, wherein said information retrieval system comprises a confidence determination subsystem for computing inter-patch confidence values between said patches and intra-patch confidence values, and said cluster estimation subsystem selects said patches depending on said inter-patch confidence values to represent clusters of said document-keyword vectors.

Claim 12. (Withdrawn) The system of claim 10, wherein said cluster estimation subsystem estimates sizes of said clusters depending on said intra-patch confidence values.

Claim 13. (Withdrawn) The system of claim 10, wherein said system further comprises a user query receiving subsystem for receiving said query and extracting data for information retrieval to generate a query vector, and an information retrieval subsystem for computing similarities between said document-keyword vectors and said query vector to select said document-keyword vectors.

Claim 14. (Withdrawn) The system of claim 10, wherein said clusters are estimated using said retrieved document-keyword vectors with respect to said user input query.

Claim 15. (Withdrawn) A graphical user interface system for graphically presenting estimated clusters on a display device in response to a user input query, said graphical user interface system comprising:

a database for storing documents;

a computer for generating document-keyword vectors for said documents stored in said database and for estimating clusters of documents in response to said user input query;
and

a display for displaying on screen said estimated clusters together with confidence relations between said clusters and hierarchical information pertaining to cluster size.

Claim 16. (Withdrawn) The graphical user interface system of claim 15, wherein said computer comprises:

a neighborhood patch generation subsystem for generating groups of nodes having similarities as determined using a search structure, said neighborhood patch generation

subsystem including a subsystem for generating a hierarchical structure upon said document-keyword vectors and a patch defining subsystem for creating patch relationships among said nodes with respect to a metric distance between nodes; and

a cluster estimation subsystem for generating cluster data of said document-keyword vectors using said similarities of patches.

Claim 17. (Withdrawn) The graphical user interface system of claim 15, wherein said computer comprises a confidence determination subsystem for computing inter-patch confidence values between said patches and intra-patch confidence values, and said cluster estimation subsystem selects said patches depending on said inter-patch confidence values to represent clusters of said document-keyword vectors and said cluster estimation subsystem estimates sizes of said clusters depending on said intra-patch confidence values.